VISVESVARAYA TECHNOLOGICAL UNIVERSITY
BELAGAVI – 590018

An Internship Report on

"THYROID DETECTION USING RANDOM FOREST"

Submitted in partial fulfillment of the requirements for the degree of

BACHELOR OF ENGINEERING

IN

INFORMATION SCIENCE AND ENGINEERING

Subject: INTERNSHIP [18CSI85]

Submitted By

SHREYA (4JK19IS053)

Internship carried out at

AgileTEC, Mangalore

Internal Guide	**External Guide**
Prof Rakesh M R	**Krishnaraj**
Assistant professor	**Managing Director**
AJIET Mangaluru	**AgileTEC Mangaluru**

DEPARTMENT OF INFORMATION SCIENCE AND ENGINEERING

A. J. INSTITUTE OF ENGINEERING & TECHNOLOGY

NH-66, KOTTARA CHOWKI, MANGALURU – 575006

2022 - 2023

A. J. INSTITUTE OF ENGINNERING & TECHNOLOGY

NH – 66, Kottara Chowki, Mangaluru - 575006

A Unit of Laxmi Memorial Education Trust (R)

(Affiliated to Visvesvaraya Technological University, Belagavi & Approved by AICTE, New Delhi)

DEPARTMENT OF INFORMATION SCIENCE AND ENGINEERING

CERTIFICATE

Certified that the Internship Report titled **"THYROID DETECTION USING RANDOM FOREST"** is carried out by Ms. **SHREYA**, USN: **4JK19IS053**, a bonafide student of A.J. Institute of Engineering & Technology, Mangaluru, at **AgileTEC, Mangalore** in partial fulfillment for the award of **Bachelor of Engineering** in **Information Science and Engineering** of **Visveswaraya Technological University, Belagavi** during the year 2022-2023. It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the Report deposited in the departmental library.

The report has been approved as it satisfies the academic requirements in respect of Internship prescribed for the said Degree.

Prof Rakesh M R
Internship Guide

Prof Rakesh M R
Internship Coordinator

Dr Suresha D
Head of the Department

Dr Shantharama Rai C
Principal

Examiners **Signature with Date**

1.

2.

AgileTEC

Office No: 4,
First Floor,
Lalbagh Commercial Complex,
Lalbagh , Mangalore-575003
0824-4260555 / 7975263562

CIN: U72200KA2015PTC084828
Ref No: HR/Edu/2022-23/007

October 14, 2022

To Whomsoever It May Concern:

This is to certify that **Ms. Shreya (USN: 4JK19IS053),** a final year student of A J Institute Of Engineering And Technology, Mangalore has successfully completed Internship on Machine Learning domain at AgileTEC, Mangaluru from August 27, 2022 to September 26, 2022 and completed a project called "**Thyroid Detection**" to fulfill the partial requirement of her B.E degree.

With Thanks and Regards,

Mr. Krishnaraj,

Managing Director,
AgileTEC,
Mangaluru.

AgileTEC
Mangaluru
Ph : 0824-4260555
A Unit of DevAppSys IT Sol Pvt. Ltd.

ACKNOWLEDGEMENT

First and foremost, we thank our parents for what we are and where we are today, without whose hard work and sacrifice we would not be here today.

Hearty thanks to **AgileTec**, for providing all the facilities that helped us in timely completion of Internship work.

We deem it a privilege to place on record the deep sense of gratitude to our internship coordinator and internship guide, **Prof. Rakesh MR, Assistant Professor,** Department of Information Science and Engineering, Department of Information Science and Engineering, who always stood behind us and supported in each and every step of the internship work.

We are grateful to **Dr. Suresha D.,** Head of the Department, Information Science and Engineering for his support and encouragement.

We are indebted to our respected Principal **Dr. Shantharama Rai. C,** beloved Vice President **Mr. Prashanth Shetty** and the management of **A. J. Institute of Engineering and Technology, Mangaluru** for providing all the facilities that helped us in timely completion of this internship report.

Finally, we would like to thank all the teaching and non-teaching staff of Department of Information Science and Engineering for their valuable help and support.

Name of the Student: Shreya
USN:4JK19IS053

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATION

Short form	**Full form**
CT	Computed Tomography
KNN	K-Nearest Neighbour
ML	Machine Learning
MRI	Magnetic Resonance Imagining
SVM	Support Vector Machine
T3	Triiodothyronine
T4	Thyroxine
TSH	Thyroid-Stimulating Hormone

CHAPTER 1

EXECUTIVE SUMMARY

This report refers to the work completed during my internship with the AgileTEC Mangalore in Machine learning using the python domain from August 27, 2022 to September 26, 2022. Thyroid detection using machine learning project was performed during internship.

Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy. IBM has a rich history with machine learning. ML is one of the most exciting technologies that one would have ever come across. It provides a collection of algorithms, methods, and tools able to embody some kind of intelligence to machines. There are four basic approaches: supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning. The type of algorithm data scientists chooses to use depends on what type of data they want to predict.

Machine learning is a method of teaching computers to learn from data, without being explicitly programmed. Python is a popular programming language for machine learning because it has a large number of powerful libraries and frameworks that make it easy to implement machine learning algorithms. To get started with machine learning using Python, you will need to have a basic understanding of Python programming and some knowledge of mathematical concepts such as probability, statistics, and linear algebra.

Amidst all the hype around big data, we keep hearing the term “Machine Learning.” Not only does it offer a remunerative career, it promises to solve problems and benefit companies by making predictions and helping them make better decisions. In this blog, we will learn the Advantages and Disadvantages of Machine Learning. As we will try to understand where to use it and where not to use Machine learning.

It has been a great journey since then and in the past few decades, the progress made in the field of machine learning is simply incomparable as we can see how a search engine like google can produce a search result within a fraction of seconds and the games have reached to the level of reality

Machine learning was a great domain to perform internship. As a result, AgileTEC, a software development company rendering web-based development and software-based development company by Mr. Krishnaraj, and prepared to explore new things in a new environment for a duration of 4 weeks.

This is a very good platform for entrepreneurs to get guidance to implement their innovative ideas. Opted for machine learning domain because Machine learning is actively being used today, perhaps in many more places than one would expect., thus machine tools enable organizations to more quickly identify profitable opportunities and potential risks.

The practical applications of machine learning drive business results which can dramatically affect a company's bottom line. New techniques in the field are evolving rapidly and expanded the application of machine learning to nearly limitless possibilities. Industries that depend on vast quantities of data- and need a system to analyze it efficiently and accurately, have embraced machine learning as the best way to build models, strategize, and plan.

Another example of the use of machine learning and how it is increasing in popularity is blockchain technology and how different cryptocurrencies are based on the technology and are considered more secure than the normal monetary transactions technology and that is why the importance of machine learning is growing, so here we are going to discuss the pros and cons of machine learning. In the beginning, explored the different frameworks such as Jupyter notebook, and solved many errors which appeared during installation as well as while performing different tasks by myself and kept a separate file for future reference.

Thus, performed a project work after completing all the theory knowledge about machine learning using python. Project topic called thyroid detection using machine learning was chosen, and to perform this particular task Random Forest algorithm was used.

During this internship, learned complete detail about Introduction to Python and python programming skills like string, list, tuples, and class. This course also taught and made understand the details of different algorithms in machine learning such as Clustering, KNN, SVM, Logistics regression. It gave a firsthand experience in a corporate environment.

This internship was a great opportunity to gain knowledge and improved practical and debugging skills.

CHAPTER 2

COMPANY PROFILE

In this chapter, the brief introduction of the company logo, its major milestones, and the best services provided by the company. In further chapter it briefs about the specific function involved in company, roles and responsibility of the company and their domain in detail.

2.1 Problem definition

The logo and brand in its entirety are a fusion of string core rooted in the spirit of values, and a strong sense of connectedness with stakeholders, shape the future together, with them.

Figure 2.1: Company logo

2.2 Company history

AgileTEC is run by Eminent Industrial Experts. The Training center houses excellent facilities for holding effective training sessions for batches of students. AgileTEC is a company which provides Industry Relevant Training, Services and Solutions in IT/ITES segment. Our focus is on bridge the gap between Corporate and Institutes by providing conventional education and the requirements of the industry. Our training spans both IT and non-IT segments and provides a variety of career-oriented courses for different segments like students, job seekers and corporate citizens. At AgileTEC, your enthusiasm to master Software, Networking, Internet, and enterprise are coupled with our state of art training. The courses of AgileTEC Technologies have been specially created to meet the demand of the industry. We have tailored our courses after in-depth research validated and approved by leading IT Giants. The entire curriculum is designed to lead to internationally recognized certification.

2.3 Vision

The objective is to expose the students to the world of technologies where they develop and improve the skills of the students in various aspects such as problem-solving, project management, logic creation, analytic abilities, and much more.

2.4 Mission

AgileTEC mission is to deliver quality knowledge transfer through innovative delivery mechanism and technologies. The AgileTEC is solely for the purposes of training and development.

2.5 Services

AgileTEC company is a rapidly growing company in the field of computer application implementation, solutions, and services. They take up the courses such as Core Java, Web Designing, Programming in C, C++, C#, Android development, Computer basics, Oracle DBA training, PHP Training, Java J2EE, Spring framework, ASP.Net, SQL Server DBA, VB.Net, Hibernate Framework, .Net Framework, AJAX training.

2.6 Roles and responsibilities

AgileTEC has an expert team that offers unique solutions. All the members of the team are professional experienced and have the depth knowledge of the technology. AgileTEC provides a customized package to suit the needs of every client and takes into consideration the needs and requirements of each client and plans different ideas to improve the client's business strategies. Every company is only good as its team. AgileTEC is fortunate enough to work with talented & creative staff. They pay special attention to each client to provide the best services. The main goal of the company is to provide the best and innovative products that will help to drive potential customers to their business.

2.7 Core Values

Experience in development, services and solutions, Knowledge of new platform and technologies, committed to work, time on time delivery of products. They work with clients to complete your needs and targets.

CHAPTER 3

PROBLEM STATEMENT AND OBJECTIVES

3.1 Problem statement

Machine learning is an increasingly popular field in computer science that involves designing and developing algorithms that can learn from and make predictions or decisions based on data. While machine learning has shown remarkable success in solving complex problems in various domains, there are still several challenges and limitations that need to be addressed. Some common problem statements in machine learning include:

1. Data quality: Machine learning algorithms heavily rely on the quality and quantity of data they are trained on. If the data is biased, incomplete, or inconsistent, it can significantly impact the accuracy and generalization of the model.
2. Overfitting and underfitting: Overfitting occurs when the model becomes too complex and fits the training data too well, resulting in poor generalization on new data. Underfitting occurs when the model is too simple and fails to capture the underlying patterns in the data.
3. Interpretability and transparency: Many machine learning models are black boxes, meaning it is challenging to understand how they arrived at their decisions or predictions. This lack of interpretability and transparency can make it difficult to trust and use these models in critical applications.
4. Scalability and efficiency: As the size and complexity of data continue to grow, machine learning algorithms need to be scalable and efficient enough to handle large-scale datasets.
5. Ethics and fairness: Machine learning models can inadvertently perpetuate biases and discrimination in the data they are trained on, leading to unfair or discriminatory outcomes. It is crucial to ensure that machine learning models are ethically and fairly designed and deployed.

The Thyroid gland is a vascular gland and one of the most important organs of a human body. This gland secretes two hormones which help in controlling the metabolism of the body. The two types of Thyroid disorders are Hyperthyroidism and Hypothyroidism. When this disorder occurs in the body, they release certain type of hormones into Thyroid related Blood test is used to detect this the body which imbalances the body’s metabolism. Thyroid related test is used to detect this disease but it is often blurred and noise will be present.

Data cleansing methods were used to make the data primitive enough for the analytics to show the risk of patients getting this disease. Machine Learning plays a very deciding role in the disease prediction. Machine Learning algorithms, SVM - support vector machine, decision tree, logistic regression, KNN - K-nearest neighbors, ANN- Artificial Neural Network are used to predict the created to get data from users to predict patient's risk of getting thyroid disease.

According to statistics, thyroid disorders are on the rise in India. Approximately 1 in 10 Indian adults suffer from thyroid problem. It has been estimated thataround 42 million peoples suffer from thyroid disease. Predicting thyroid disorder by doctor is a tedious process which might lead to negative prediction, onlyexperienced doctor can examine the case properly. Toassist doctors machine learning can help them in diagnosis of disease and reduces their burden.

The problem statement of thyroid detection using machine learning is to develop an accurate and efficient algorithm that can automatically classify thyroid nodules as either malignant or benign based on medical imaging data such as ultrasound, CT, or MRI.

Thyroid cancer is one of the most common types of cancer, and early detection is crucial for successful treatment. Medical imaging techniques are widely used to detect and diagnose thyroid nodules, but the interpretation of these images is highly dependent on the experience and expertise of the radiologist.

Random Forest algorithm can help overcome this limitation by analyzing large amounts of medical imaging data to identify patterns and features that are indicative of thyroid cancer. By training a machine learning model on a large dataset of labeled thyroid images, it is possible to create a highly accurate and efficient classifier that can assist radiologists in their diagnosis and improve patient outcomes.

3.2 Objectives

The main objective of using a random forest algorithm for thyroid detection is to accurately classify thyroid diseases based on a set of input features.

1) Accurate Diagnosis: The primary objective of using the Random Forest algorithm for thyroid detection is to achieve accurate diagnosis. Random Forest is a machine learning algorithm that can handle complex data and generate accurate predictions. By analyzing patient data, the algorithm can identify patterns and make predictions about whether a patient has thyroid dysfunction or not.
2) Improved Patient Outcomes: By accurately detecting and diagnosing thyroid dysfunction and providing personalized treatment recommendations, the Random

Forest algorithm can help improve patient outcomes. This can result in better quality of life, improved health outcomes, and reduced healthcare costs.

Overall, the Random Forest algorithm can help healthcare professionals make informed decisions about the diagnosis, treatment, and management of thyroid dysfunction, resulting in improved patient outcomes.

- The main objective is to develop a system which can predict whether the patient has symptoms of thyroid or no.
- It finds the accuracy of the model used i.e., Random Forest Algorithm.

CHAPTER 4

WEEKLY OVERVIEW

Table 4.1: Weekly overview of internship

	Date	**Day**	**Task/ Topic Completed**
Week - I	27-8-2022	Sat	Company introduction and orientation about python and project
	29-8-2022	Mon	Introduction to Python
	30-8-2022	Tue	Demo project about python
	31-8-2022	Wed	Basic python (string, list, tuple, class)
	01-9-2022	Thu	Project – Installing jupyter notebook
	02-2-2022	Fri	Project- About the dataset, Importing dataset

	Date	**Day**	**Task/ Topic Completed**
Week - II	03-8-2022	Sat	Introduction to machine learning
	05-8-2022	Mon	Pandas
	06-8-2022	Tue	NumPy
	07-8-2022	Wed	Algorithms
	08-8-2022	Thu	Project-Features of machine learning
	09-8-2022	Fri	Project-Plotting graphs

	Date	**Day**	**Task/ Topic Completed**
Week - III	10-8-2022	Sat	Algorithm in ML and classification
	12-8-2022	Mon	Clustering
	13-8-2022	Tue	KNN, SVM
	14-8-2022	Wed	Logistic regression
	15-8-2022	Thu	Project-Training and testing of data
	16-8-2022	Fri	Project-Random Forest training on project

	Date	**Day**	**Task/ Topic Completed**
Week - IV	17-8-2022	Sat	Decision tree
	19-8-2022	Mon	Random forest
	20-8-2022	Tue	Matplotlib, seaborn
	21-8-2022	Wed	Project-Find accuracy, precision etc
	22-8-2022	Thu	Project-Confusion matrix using random forest
	23-8-2022	Fri	Overall discussion

CHAPTER 5

TASK PERFORMED

This chapter is about the tasks performed during the internship and the knowledge gained in 4 weeks of internship. It also says about the technical and non-technical skills developed and gives a detailed description of tasks performed and work done each day during my internship at AgileTEC.

The task performed in this internship includes learning the basics of python, python libraries such as NumPy, pandas, matplotlib, Machine learning basics, Machine learning techniques, different algorithms and working on project.

5.1 Phase-1

Installation of the jupyter notebook and launching of the terminal can be learnt in this section.

5.1.1 Installing Jupyter Notebook

To install Jupyter Notebook, you can follow these steps:

1. Install Python: If you don't have Python installed on your system. Choose the appropriate version of Python based on your operating system.
2. Install Jupyter Notebook: Once you have installed Python, you can install Jupyter Notebook by running the following command in your terminal or command prompt:

   ```
   pip install jupyter
   ```

   This command will download and install Jupyter Notebook and its dependencies.
3. Launch Jupyter Notebook: To launch Jupyter Notebook, simply run the following command in your terminal or command prompt:

   ```
   jupyter notebook
   ```

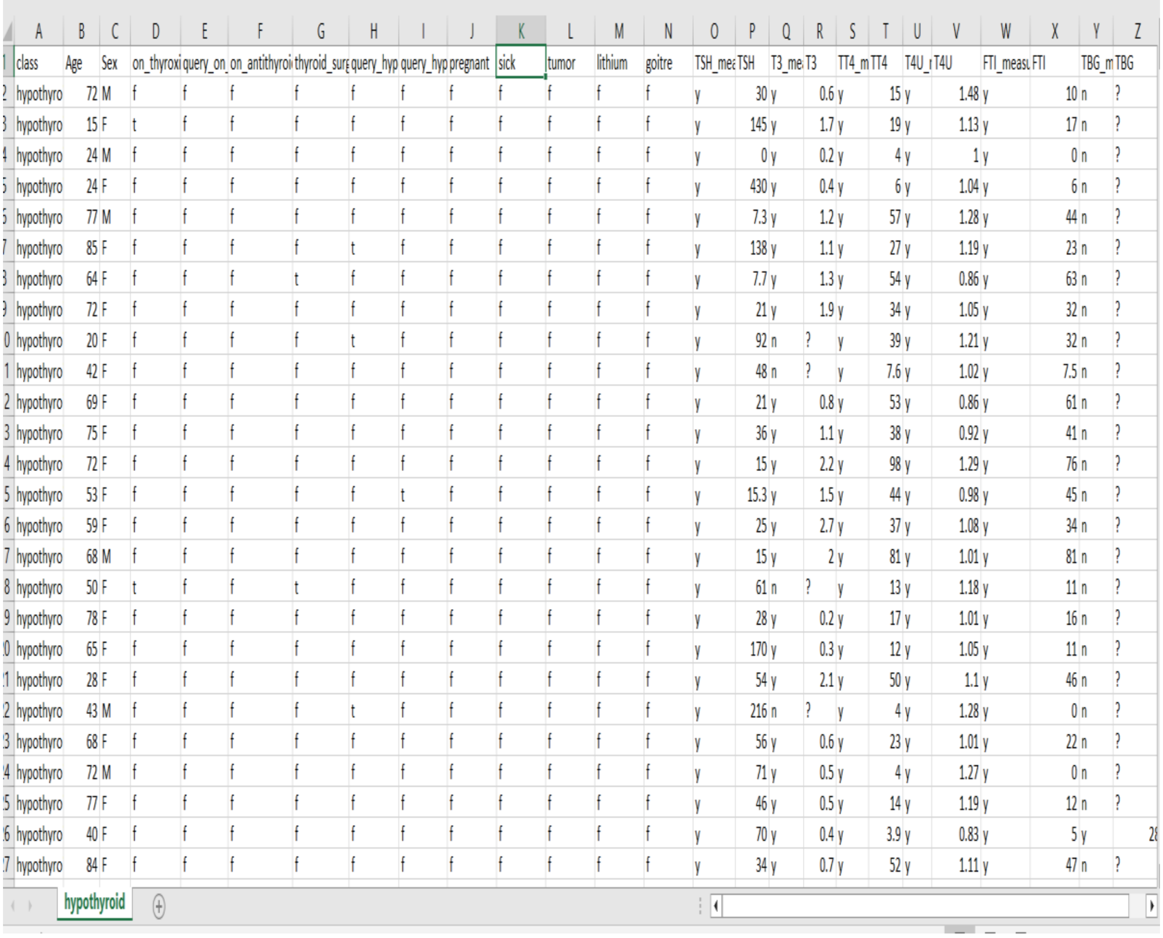
   This will start the Jupyter Notebook server and open your default web browser to the Jupyter Notebook interface. Now its able to create new Jupyter Notebook files and start working with Python in a notebook environment.

5.2 Phase-2

This section gives detailed description about the dataset, tools and instruments used for Thyroid detection using machine learning.

5.2.1 About the dataset

This project is done using dataset called “hypothyroid.csv” which is available on Kaggle contains data on the people having thyroid problem. The dataset has a total of 3138 rows and 26 columns. The dataset contains different values such as class, age, sex, on_thyroxine, query_on_thyroxine, on_antithyroid_medication, thyroid_surgery, query_hypothyroid, query_hyperthyroid, pregnant, sick, tumor, lithium, goitre, TSH_measured, TSH, T3_measured, T3, TT4_measured, TT4, T4U_measured, T4U, ETI_measured, FT1, TBG_measured, TBG. This dataset t is cleaned first by preprocessing, trained using Random Forest Algorithm and tested. Finally, prediction is done. The dataset is used for the Thyroid detection.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	class	Age	Sex	on_thyroxi	query_on_	on_antithyroi	thyroid_surg	query_hyp	query_hyp	pregnant	sick	tumor	lithium	goitre	TSH_mea	TSH	T3_me	T3	TT4_m	TT4	T4U_	T4U	FTI_measu	FTI	TBG_m	TBG
2	hypothyro	72	M	f	f	f	f	f	f	f	f	f	f	f	y	30	y	0.6	y	15	y	1.48	y	10	n	?
3	hypothyro	15	F	t	f	f	f	f	f	f	f	f	f	f	y	145	y	1.7	y	19	y	1.13	y	17	n	?
4	hypothyro	24	M	f	f	f	f	f	f	f	f	f	f	f	y	0	y	0.2	y	4	y	1	y	0	n	?
5	hypothyro	24	F	f	f	f	f	f	f	f	f	f	f	f	y	430	y	0.4	y	6	y	1.04	y	6	n	?
6	hypothyro	77	M	f	f	f	f	f	f	f	f	f	f	f	y	7.3	y	1.2	y	57	y	1.28	y	44	n	?
7	hypothyro	85	F	f	f	f	f	t	f	f	f	f	f	f	y	138	y	1.1	y	27	y	1.19	y	23	n	?
8	hypothyro	64	F	f	f	f	t	f	f	f	f	f	f	f	y	7.7	y	1.3	y	54	y	0.86	y	63	n	?
9	hypothyro	72	F	f	f	f	f	f	f	f	f	f	f	f	y	21	y	1.9	y	34	y	1.05	y	32	n	?
10	hypothyro	20	F	f	f	f	f	t	f	f	f	f	f	f	y	92	n	?	y	39	y	1.21	y	32	n	?
11	hypothyro	42	F	f	f	f	f	f	f	f	f	f	f	f	y	48	n	?	y	7.6	y	1.02	y	7.5	n	?
12	hypothyro	69	F	f	f	f	f	f	f	f	f	f	f	f	y	21	y	0.8	y	53	y	0.86	y	61	n	?
13	hypothyro	75	F	f	f	f	f	f	f	f	f	f	f	f	y	36	y	1.1	y	38	y	0.92	y	41	n	?
14	hypothyro	72	F	f	f	f	f	f	f	f	f	f	f	f	y	15	y	2.2	y	98	y	1.29	y	76	n	?
15	hypothyro	53	F	f	f	f	f	f	t	f	f	f	f	f	y	15.3	y	1.5	y	44	y	0.98	y	45	n	?
16	hypothyro	59	F	f	f	f	f	f	f	f	f	f	f	f	y	25	y	2.7	y	37	y	1.08	y	34	n	?
17	hypothyro	68	M	f	f	f	f	f	f	f	f	f	f	f	y	15	y	2	y	81	y	1.01	y	81	n	?
18	hypothyro	50	F	t	f	f	t	f	f	f	f	f	f	f	y	61	n	?	y	13	y	1.18	y	11	n	?
19	hypothyro	78	F	f	f	f	f	f	f	f	f	f	f	f	y	28	y	0.2	y	17	y	1.01	y	16	n	?
20	hypothyro	65	F	f	f	f	f	f	f	f	f	f	f	f	y	170	y	0.3	y	12	y	1.05	y	11	n	?
21	hypothyro	28	F	f	f	f	f	f	f	f	f	f	f	f	y	54	y	2.1	y	50	y	1.1	y	46	n	?
22	hypothyro	43	M	f	f	f	f	t	f	f	f	f	f	f	y	216	n	?	y	4	y	1.28	y	0	n	?
23	hypothyro	68	F	f	f	f	f	f	f	f	f	f	f	f	y	56	y	0.6	y	23	y	1.01	y	22	n	?
24	hypothyro	72	M	f	f	f	f	f	f	f	f	f	f	f	y	71	y	0.5	y	4	y	1.27	y	0	n	?
25	hypothyro	77	F	f	f	f	f	f	f	f	f	f	f	f	y	46	y	0.5	y	14	y	1.19	y	12	n	?
26	hypothyro	40	F	f	f	f	f	f	f	f	f	f	f	f	y	70	y	0.4	y	3.9	y	0.83	y	5	y	28
27	hypothyro	84	F	f	f	f	f	f	f	f	f	f	f	f	y	34	y	0.7	y	52	y	1.11	y	47	n	?

Figure 5.1 hypothyroid.csv dataset

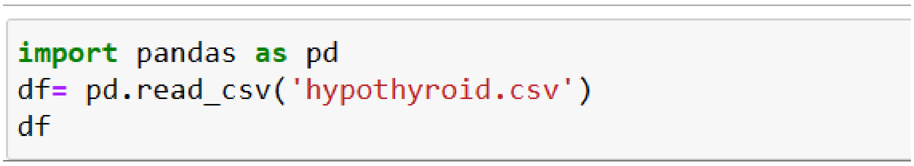
The features in the dataset are:

1. Class: It consists of two classes hypothyroid or negative. Where hypothyroid indicates patients having thyroid problem, negative indicates patient does not have thyroid problem.
2. Age: The age of the patient.
3. Sex: The gender of the patient.
4. on_thyroxine: Whether patient is on thyroxine
5. query on thyroxine: Whether patient is on thyroxine

6. on antithyroid meds: Whether the patient is on antithyroid meds.
7. sick: Whether patient is sick
8. pregnant: Whether patient is pregnant
9. thyroid_surgery: Whether patient has undergone thyroid surgery
10. query_hypothyroid: Whether the patient believes they have hypothyroid
11. query_hyperthyroid: Whether the patient believes they have hyperthyroid
12. lithium: Whether patient has lithium
13. goitre: Whether patient has goitre
14. tumor: Whether patient has tumor
15. TSH_measured: Whether TSH was measured in the blood
16. TSH: TSH level in blood from lab work
17. T3_measured: Whether T3 was measured in the blood
18. T3: T3 level in blood from lab work
19. TT4_measured: Whether TT4 was measured in the blood
20. TT4: TT4 level in blood from lab work
21. T4U_measured: Whether T4U was measured in the blood
22. T4U: T4U level in blood from lab work
23. FTI_measured: Whether FTI was measured in the blood
24. FTI: FTI level in blood from lab work
25. TBG_measured: Whether TBG was measured in the blood
26. TBG: TBG level in blood from lab work

5.2.2 Importing Dataset

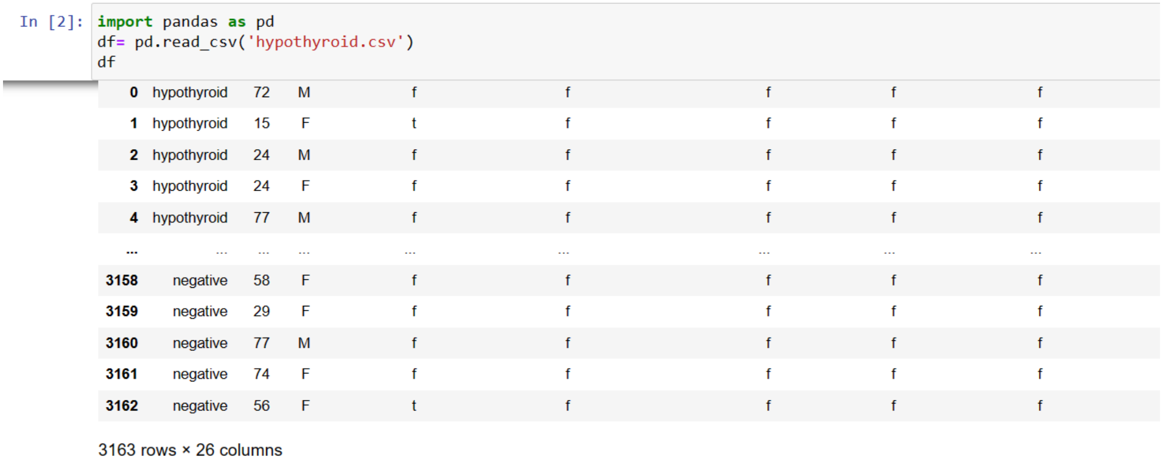
Importing datasets using Pandas is essential for analyzing and visualizing data in various research projects. In this case, thyroid detection dataset can be imported using Pandas in several steps. First, the data can be downloaded from a reliable source and saved in a CSV file format. Then, using the Pandas library, the CSV file can be read and loaded into a Data Frame object.



```
import pandas as pd
df= pd.read_csv('hypothyroid.csv')
df
```

Figure 5.2 Importing dataset

The obtained result after importing the dataset is shown below:



```
In [2]: import pandas as pd
        df= pd.read_csv('hypothyroid.csv')
        df
```

0	hypothyroid	72	M	f	f	f	f	f
1	hypothyroid	15	F	t	f	f	f	f
2	hypothyroid	24	M	f	f	f	f	f
3	hypothyroid	24	F	f	f	f	f	f
4	hypothyroid	77	M	f	f	f	f	f
...	...	...	...	...	...	...	...	...
3158	negative	58	F	f	f	f	f	f
3159	negative	29	F	f	f	f	f	f
3160	negative	77	M	f	f	f	f	f
3161	negative	74	F	f	f	f	f	f
3162	negative	56	F	t	f	f	f	f

3163 rows × 26 columns

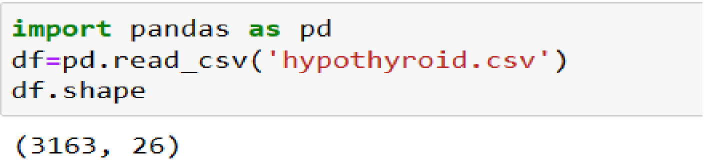
Figure 5.3: Imported dataset result

5.3 Phase-3

In this phase, different features in the machine learning and how it is used in our project has been shown.

5.3.1 Shape of the dataset

In machine learning, "shape" refers to the dimensions or structure of a data object, such as a tensor or array. The shape is defined by the number of columns and number of rows in the dataset. So, in this project there are 3163 rows and 26 columns. Technically this is a 2-dimensional array.



```
import pandas as pd
df=pd.read_csv('hypothyroid.csv')
df.shape
```

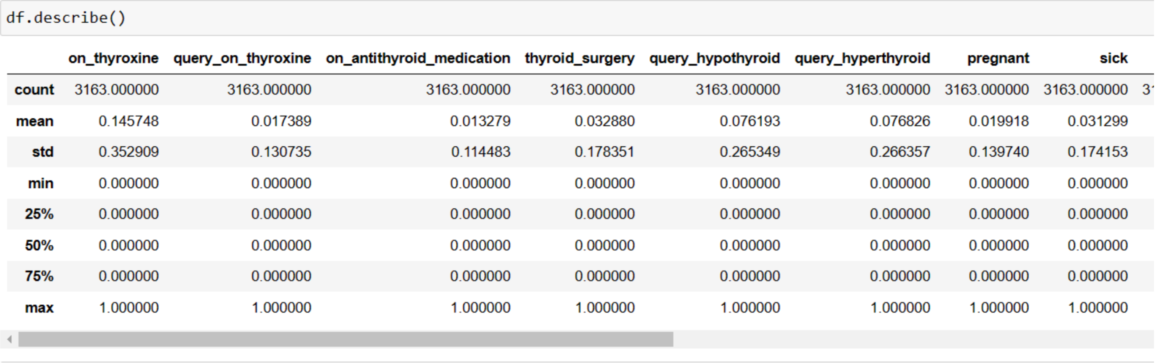
```
(3163, 26)
```

Figure 5.4: Shape of the dataset

5.3.2 Attribute describe ()

Basically, describe () method will return some basic information about each column in the dataset. In Python, describe () is a built-in method of the pandas library, which is commonly used in data analysis and manipulation. It calculates several statistical measures of central tendency, dispersion, and shape of the distribution of a numeric dataset, including the mean, standard deviation, minimum, maximum, and quartile ranges.

The output of describe () provides valuable insights into the dataset, such as identifying any potential outliers, determining the distribution of the data, and identifying any patterns or trends that may exist.



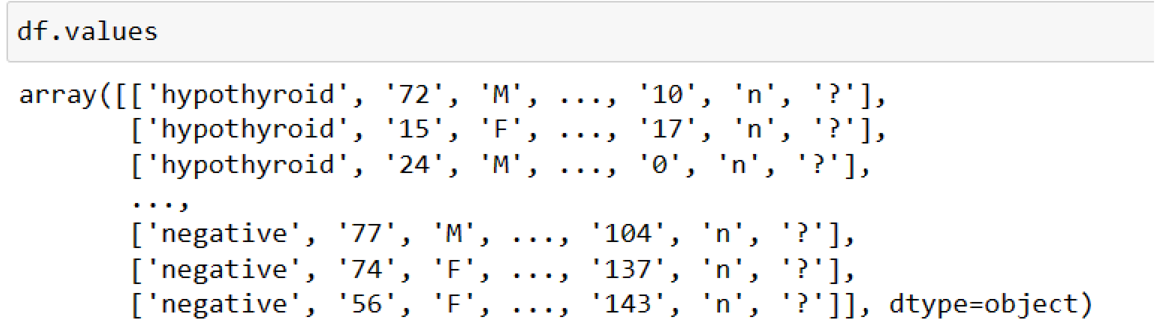
```
df.describe()
```

	on_thyroxine	query_on_thyroxine	on_antithyroid_medication	thyroid_surgery	query_hypothyroid	query_hyperthyroid	pregnant	sick
count	3163.000000	3163.000000	3163.000000	3163.000000	3163.000000	3163.000000	3163.000000	3163.000000
mean	0.145748	0.017389	0.013279	0.032880	0.076193	0.076826	0.019918	0.031299
std	0.352909	0.130735	0.114483	0.178351	0.265349	0.266357	0.139740	0.174153
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
50%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
75%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
max	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000

Figure 5.5: Attribute describe() result

5.3.3 Attribute value()

df.value returns a numpy array with the underlying data of the Data Frame, without any index or columns names.



```
df.values
```

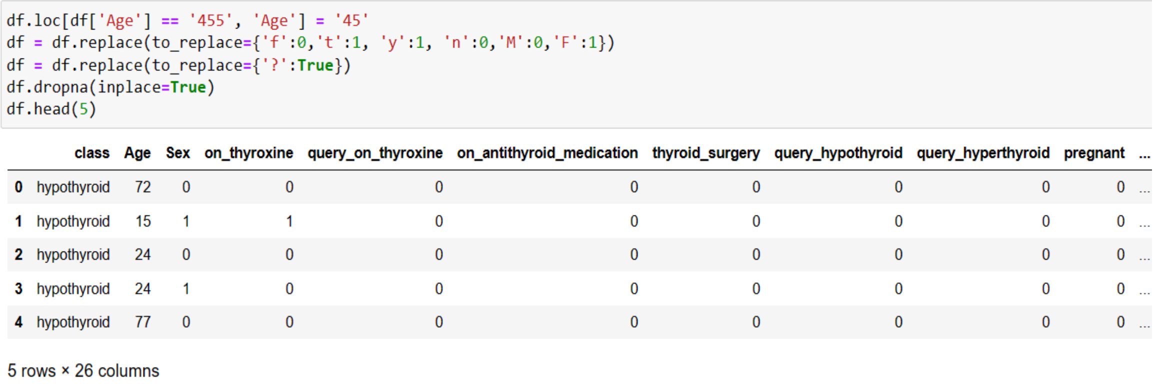
```
array([['hypothyroid', '72', 'M', ..., '10', 'n', '?'],
       ['hypothyroid', '15', 'F', ..., '17', 'n', '?'],
       ['hypothyroid', '24', 'M', ..., '0', 'n', '?'],
       ...,
       ['negative', '77', 'M', ..., '104', 'n', '?'],
       ['negative', '74', 'F', ..., '137', 'n', '?'],
       ['negative', '56', 'F', ..., '143', 'n', '?']], dtype=object)
```

Figure 5.6 Attribute value () result

5.3.4 Replacing with numeric value

The replace () method is a built-in functionality offered in Python programming. It replaces all the occurrences of the old substring with the new substring. Replace () returns a new string in which old substring is replaced with the new substring. So here string is replaced with the numeric value.

Here, in the Sex column M is replaced with 0 and F is replaced with 1. And f, n is replaced by 0 and t, y is replaced by 1. Then the symbol ‘?’ is replaced by True.



```
df.loc[df['Age'] == '455', 'Age'] = '45'
df = df.replace(to_replace={'f':0,'t':1, 'y':1, 'n':0,'M':0,'F':1})
df = df.replace(to_replace={'?':True})
df.dropna(inplace=True)
df.head(5)
```

	class	Age	Sex	on_thyroxine	query_on_thyroxine	on_antithyroid_medication	thyroid_surgery	query_hypothyroid	query_hyperthyroid	pregnant	...
0	hypothyroid	72	0	0	0	0	0	0	0	0	...
1	hypothyroid	15	1	1	0	0	0	0	0	0	...
2	hypothyroid	24	0	0	0	0	0	0	0	0	...
3	hypothyroid	24	1	0	0	0	0	0	0	0	...
4	hypothyroid	77	0	0	0	0	0	0	0	0	...

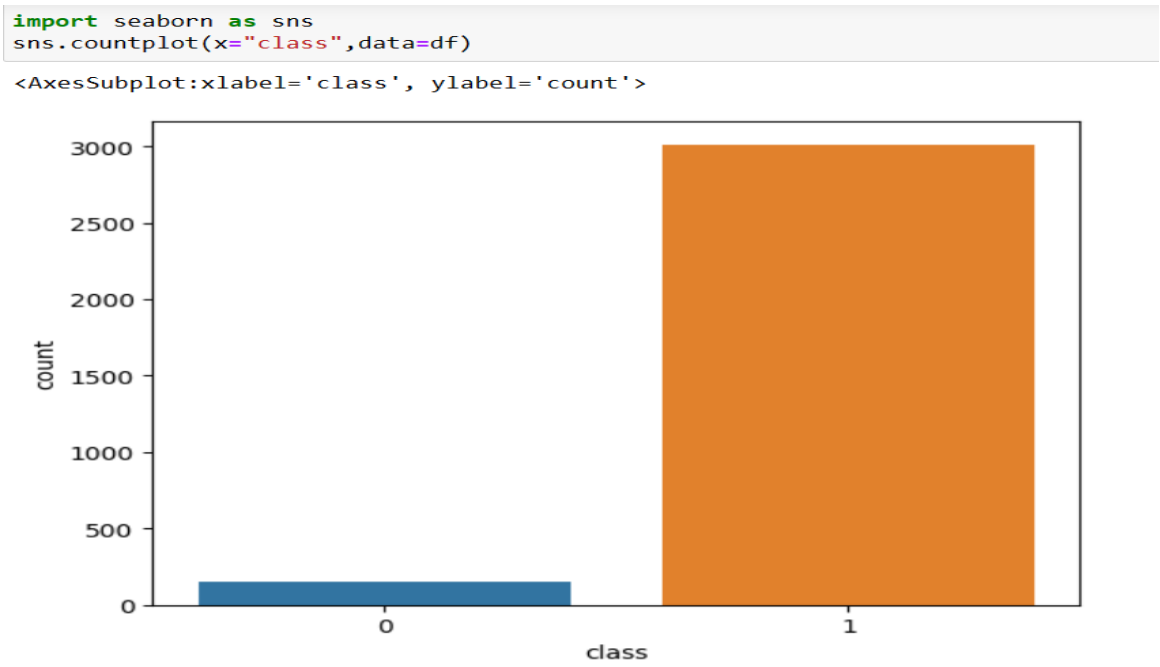
5 rows × 26 columns

Figure 5.7 String replace with numeric value

5.3.5 Plotting countplot

The countplot () function for visualizing data of your machine learning or data science project. The count plot is majorly used for showing the observational count in different category-based bins with the help of bars. Show the counts of observations in each categorical bin using bars. A count plot can be thought of as a histogram across a categorical, instead of quantitative, variable. Here, xlabel shows the class and ylabel shows the count.

In the below graph we can see '0' indicates the hypothyroid class whereas '1' indicates the negative class. Thus, the number of negative patients is more compared to patients having hypothyroid.



```
import seaborn as sns
sns.countplot(x="class",data=df)
```

```
<AxesSubplot:xlabel='class', ylabel='count'>
```

Figure 5.8: Plotting countplot()

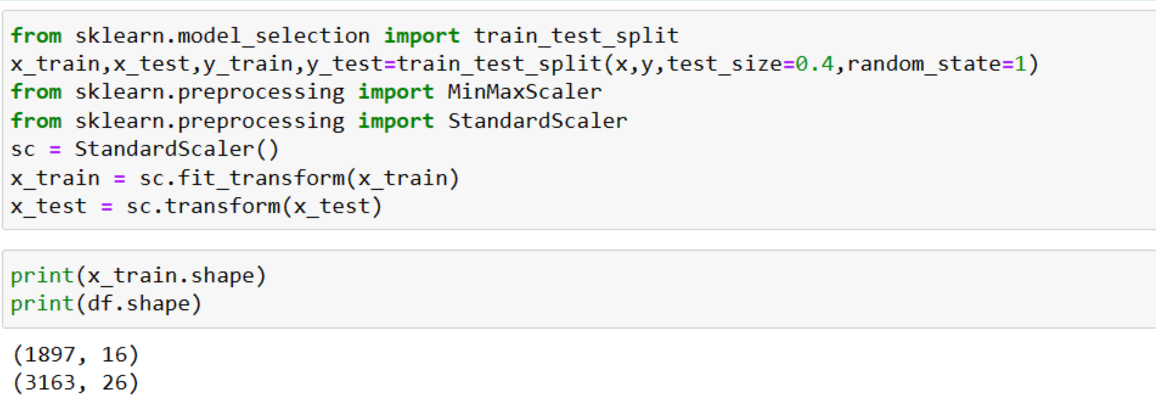
5.4 Phase-4

In this section, training and testing of data is performed. Random forest classifier is used for training and testing of dataset.

5.4.1 Training and Testing the data

Training data split is a crucial step in machine learning and data analysis projects as it allows for the evaluation of model performance on an independent dataset. By splitting the thyroid dataset into training and testing subsets, the model can be trained on the training set and tested on the testing set to evaluate its accuracy and effectiveness. This helps to avoid overfitting, which occurs when the model learns from the training set but fails to generalize to new data. Including the results of the training data split in the report provides stakeholders with insights into the model's performance and its ability to detect thyroid cases accurately.

Firstly, we will import train_test_split model for training and testing of data and put labels for the values. Here ,test_size is taken as 0.4 which means 60% is taken for training and 40% for testing. For the hypothyroid.csv dataset out of 3163 rows and 26 columns, 1897 rows and 16 columns are used for training and rest is used for testing.



```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.4,random_state=1)
from sklearn.preprocessing import MinMaxScaler
from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
x_train = sc.fit_transform(x_train)
x_test = sc.transform(x_test)
```

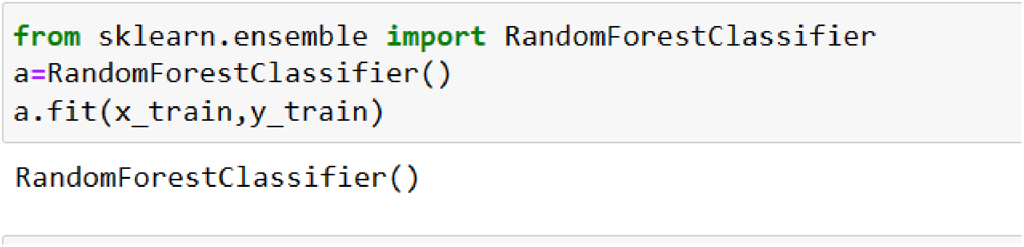
```
print(x_train.shape)
print(df.shape)
```

```
(1897, 16)
(3163, 26)
```

Figure 5.9 Training and testing dataset

5.4.2 Import Random Forest Classifier

The RandomForestClassifier library is imported to train the data. x_train and y_train are fitted into random forest classifier.



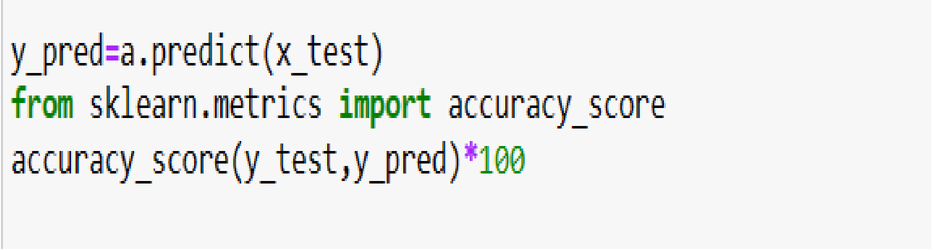
```
from sklearn.ensemble import RandomForestClassifier
a=RandomForestClassifier()
a.fit(x_train,y_train)
```

```
RandomForestClassifier()
```

Figure 5.10 Import Random Forest classifier

5.4.3 Test Result

The accuracy of the Thyroid cases using random forest is shown below.



```
y_pred=a.predict(x_test)
from sklearn.metrics import accuracy_score
accuracy_score(y_test,y_pred)*100
```

Figure 5.11 Accuracy Score

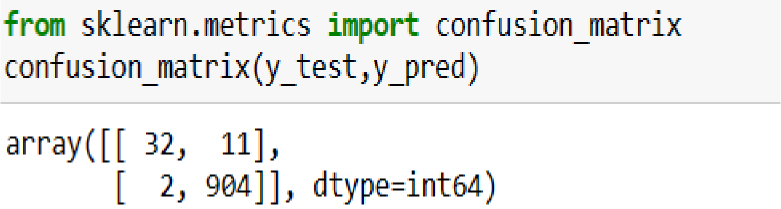
5.5 Phase-5

This phase Includes evaluation metrices performed using confusion matrix in random forest to find accuracy, precision, specificity and sensitivity.

5.5.1 Confusion matrix using random forest

A confusion matrix is a table that is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known. The confusion matrix itself is relatively simple to understand, but the related terminology can be confusing.

Confusion matrix is used to evaluate the model. A confusion matrix summarizes correct and incorrect predictions, which helps us calculate accuracy, precision, sensitivity, recall, and f1-score.



```
from sklearn.metrics import confusion_matrix
confusion_matrix(y_test,y_pred)
```

```
array([[ 32,  11],
       [  2, 904]], dtype=int64)
```

Figure 5.12 Confusion matrix using random forest

5.5.2 Calculate precision

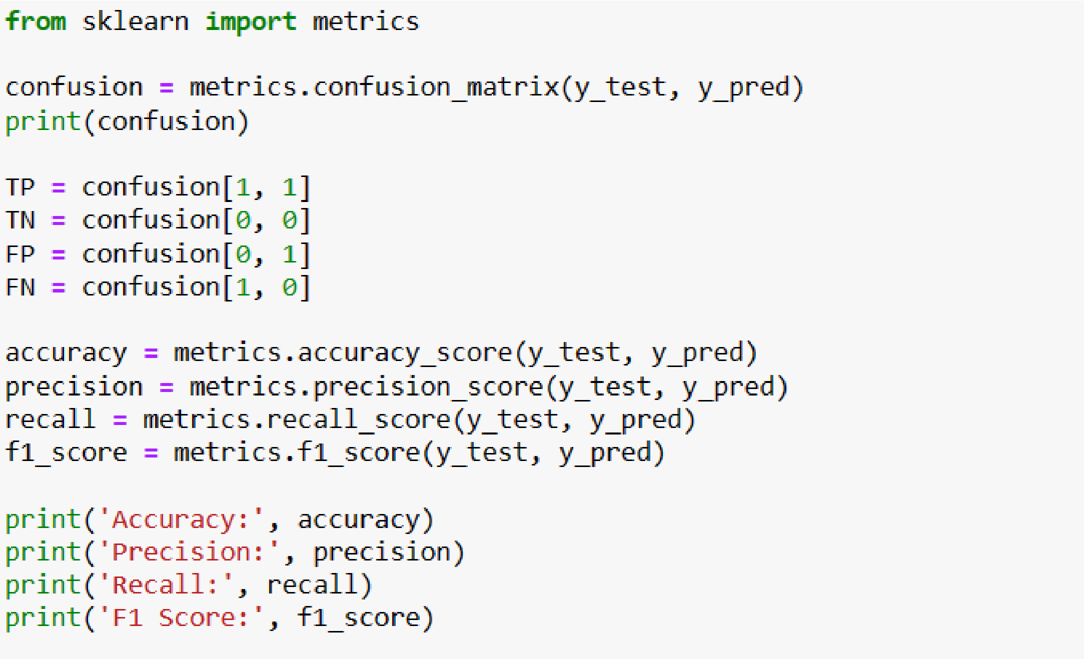
Precision is a metric used to evaluate the performance of a binary classification model. The quality of a positive prediction made by the model. Here built-in function precision_score is used to calculate the precision.

5.5.3 Calculate Recall

Recall is also called as true positive rate or sensitivity. Recall in Machine Learning can be described as the metric used for evaluating a model’s ability to predict the true positives of each available category. It is the probability of a positive test result, conditioned on the individual truly being positive.

5.5.4 Calculate f1 score

F1 score is a machine learning evaluation metric that measures a model's accuracy. It combines the precision and recall scores of a model. The accuracy metric computes how many times a model made a correct prediction across the entire dataset.



```
from sklearn import metrics

confusion = metrics.confusion_matrix(y_test, y_pred)
print(confusion)

TP = confusion[1, 1]
TN = confusion[0, 0]
FP = confusion[0, 1]
FN = confusion[1, 0]

accuracy = metrics.accuracy_score(y_test, y_pred)
precision = metrics.precision_score(y_test, y_pred)
recall = metrics.recall_score(y_test, y_pred)
f1_score = metrics.f1_score(y_test, y_pred)

print('Accuracy:', accuracy)
print('Precision:', precision)
print('Recall:', recall)
print('F1 Score:', f1_score)
```

Figure 5.13 Calculating Metrics

CHAPTER 6

METHODOLOGY

This chapter gives detailed description about the project, its need, system architecture, implementation of the project Thyroid detection using random forest.

Methodology in a project refers to the set of procedures, techniques, and tools that are used to plan, execute, monitor, and control the project from start to finish. It is a structured approach that helps ensure that the project is completed on time, within budget, and meets the required quality standards.

Methodology can be applied to a wide range of fields, including scientific research, business management, education, social sciences, and technology development. In each field, different methodologies may be used depending on the specific goals and objectives of the project or study.

Methodology is important because it helps to ensure that processes are carried out in a standardized way, which makes it easier to measure and compare results. It also provides a framework for analyzing and understanding complex problems, as well as for identifying and implementing effective solutions.

6.1 About the project

Thyroid detection using a random forest algorithm is a machine learning project that aims to predict whether a patient has thyroid disease based on various features such as age, sex, TSH levels, T3 levels, and T4 levels.

Random forest is a popular machine learning algorithm used for classification tasks. It is an ensemble learning method that builds multiple decision trees and combines their predictions to produce a final prediction. In this case, the decision trees are built using the input features to predict whether a patient has thyroid disease or not.

To develop such a project, you would need a dataset of thyroid patients, which includes the features mentioned above and their corresponding thyroid disease status. You would then split the dataset into training and testing sets and train a random forest model on the training set. After training the model, you would evaluate its performance on the testing set to assess its accuracy.

If the performance of the model is satisfactory, you can use it to predict thyroid disease status for new patients based on their input features. However, it is important to note that this model should not be used as a substitute for professional medical diagnosis or advice. Its purpose is to assist medical professionals in their diagnosis by providing additional information based on the input features.

6.1.1 Advantages of Thyroid detection using random forest

Using a random forest algorithm for thyroid detection can have several advantages. Here are a few reasons why it may be useful:

- Accurate classification: Random Forest algorithm has been shown to provide accurate results in classification problems. It can effectively classify patients into different categories, such as normal, hyperthyroidism, hypothyroidism, and thyroid nodules.
- Robustness: Random Forest is a robust algorithm that can handle missing data, noisy data, and outliers. This is important in medical diagnosis where the data can be incomplete or noisy.
- Feature importance: Random Forest algorithm can also provide information about the importance of the input features used for classification. This can be helpful in identifying the most significant features for thyroid detection.
- Scalability: Random Forest algorithm is a scalable algorithm that can handle large datasets efficiently. This is important in medical diagnosis where the amount of data can be substantial.

In summary, using a random forest algorithm for thyroid detection can provide accurate results, handle missing or noisy data, identify significant features, and handle large datasets efficiently. Therefore, it can be a useful tool in medical diagnosis and research related to thyroid disorders.

6.1.2 System Architecture

System architecture refers to the design and organization of a computer or software system, including its components, structure, and behavior. It involves determining the system's functionality, performance, scalability, reliability, security, and other characteristics, as well as how the system will interact with other systems or components.

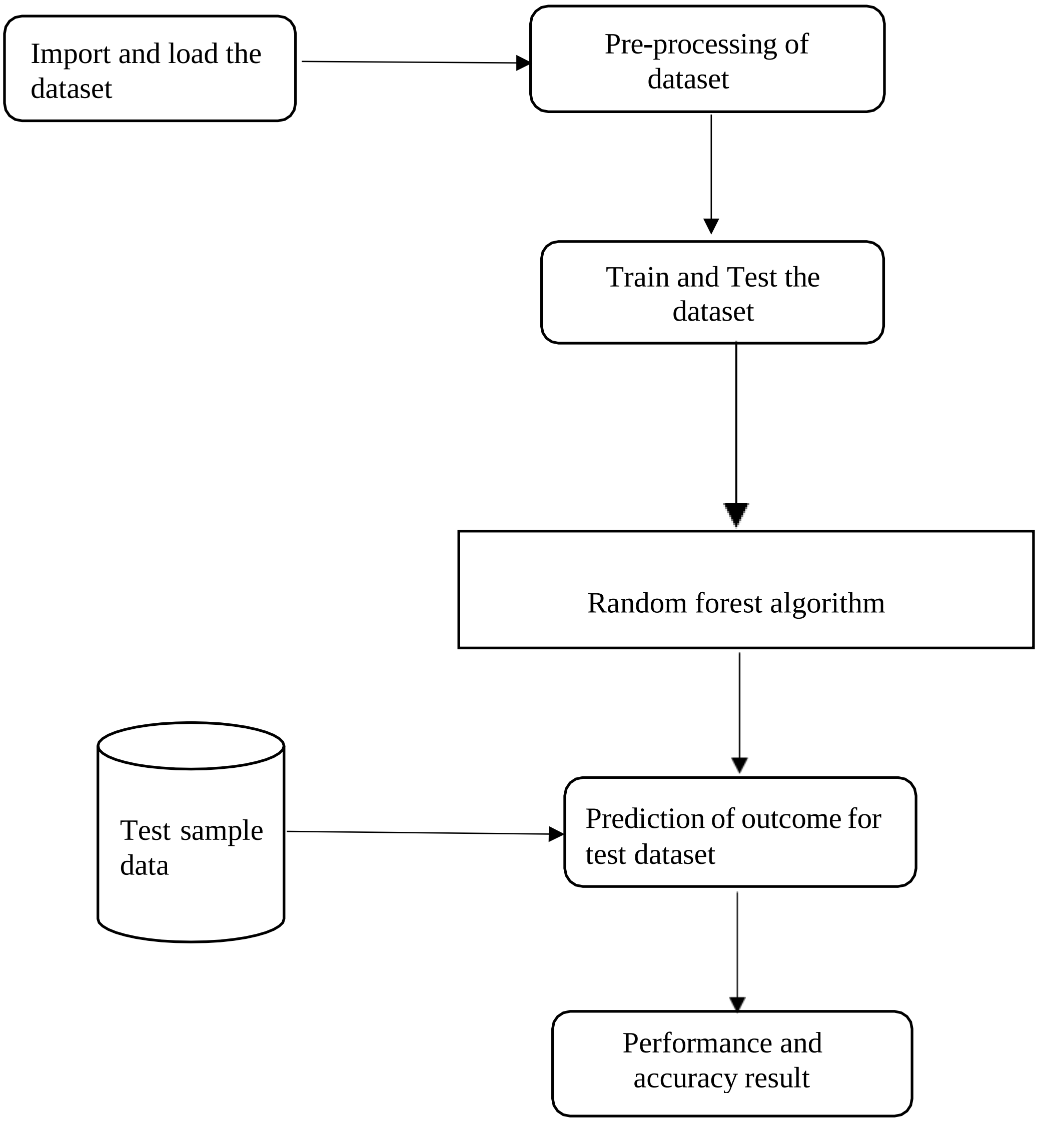

Import and load the dataset
Pre-processing of dataset
Train and Test the dataset
Random forest algorithm
Test sample data
Prediction of outcome for test dataset
Performance and accuracy result

Figure 6.1 System Architecture

6.2 Implementation

Thyroid detection using the random forest algorithm involves several steps:

6.2.1 Data Collection

Thyroid detection involves the collection and analysis of data from various sources in order to diagnose or screen for thyroid disorders.

First, a dataset containing features related to thyroid disorders such as age, gender, TSH levels, T3 levels, T4 levels, etc. needs to be collected.

The data collection step in thyroid detection involves gathering information from various sources to help diagnose or screen for thyroid disorders. The specific tests and procedures performed will depend on the individual patient's symptoms and medical history.

6.2.2 Data Preprocessing

Data preprocessing is an essential step in building a machine learning model for thyroid detection using random forest. The goal of data preprocessing is to clean, transform, and prepare the data for modeling.

Here are the steps involved in data preprocessing for thyroid detection using random forest:

- Data Splitting: Split the data into training and testing sets to evaluate the model's performance. The training set is used to train the model, and the testing set is used to evaluate the model's performance on unseen data.
- Model Building: Train a random forest model on the training data.
- Model Evaluation: Evaluate the model's performance on the testing data using various metrics such as accuracy, precision, recall, sensitivity, selectivity and etc.

In conclusion, data preprocessing is a crucial step in building a machine learning model for thyroid detection using random forest. It involves cleaning, transforming, and preparing the data for modeling, which ultimately improves the model's accuracy and reliability.

6.2.3 Random Forest Training

Random forest is a popular machine learning algorithm that is commonly used for classification tasks Once the relevant features are selected, a random forest model is trained using the training dataset. The random forest algorithm involves building multiple decision trees and then combining their predictions to make a final prediction. In the context of thyroid detection, random forest can be trained on a dataset along with corresponding labels indicating the presence or absence of a thyroid nodule.

6.2.4 Model Evaluation

When evaluating a thyroid detection model using the Random Forest algorithm, there are several metrics that can be used to assess its performance.

Here are some commonly used metrics:

- Accuracy
- Precision
- Recall
- F1 score

When evaluating a thyroid detection model, it is important to consider the specific goals of the application, as different metrics may be more or less important depending on the context. For example, in a screening context where the goal is to identify as many cases as possible, recall may be a more important metric than precision.

6.2.5 Deployment

Once the model has been evaluated and shown to be accurate, it can be deployed for use in thyroid detection. This may involve integrating the model into a larger software system or application, or simply making it available for doctors or other medical professionals to use directly.

It is important to note that deployment of a machine learning algorithm in a medical context requires careful consideration of ethical and legal implications, as well as rigorous testing and validation to ensure patient safety and accuracy of results.

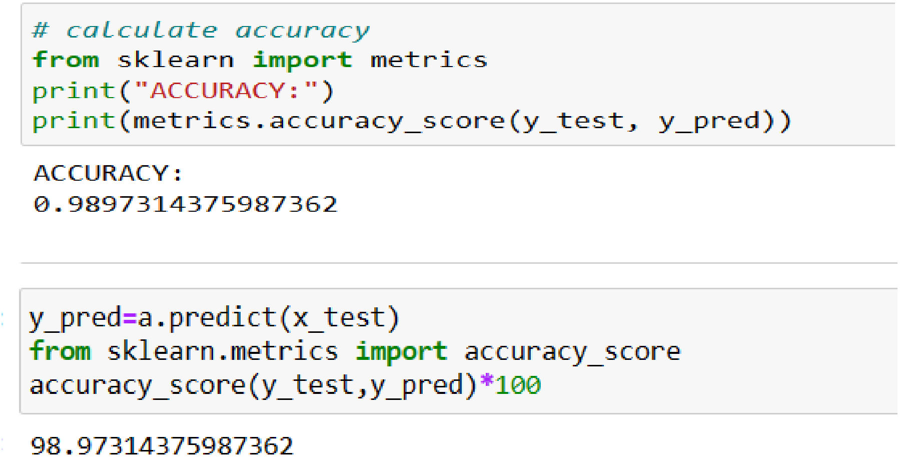
Overall, the random forest algorithm can be an effective method for thyroid detection, as it can handle non-linear relationships between the input features and the output variable and is less prone to overfitting compared to other machine learning algorithms.

CHAPTER 7

RESULT ANAYSIS

7.1 Accuracy

This section shows the result fetched by Random Forest:



```
# calculate accuracy
from sklearn import metrics
print("ACCURACY:")
print(metrics.accuracy_score(y_test, y_pred))
```

```
ACCURACY:
0.9897314375987362
```

```
y_pred=a.predict(x_test)
from sklearn.metrics import accuracy_score
accuracy_score(y_test,y_pred)*100
```

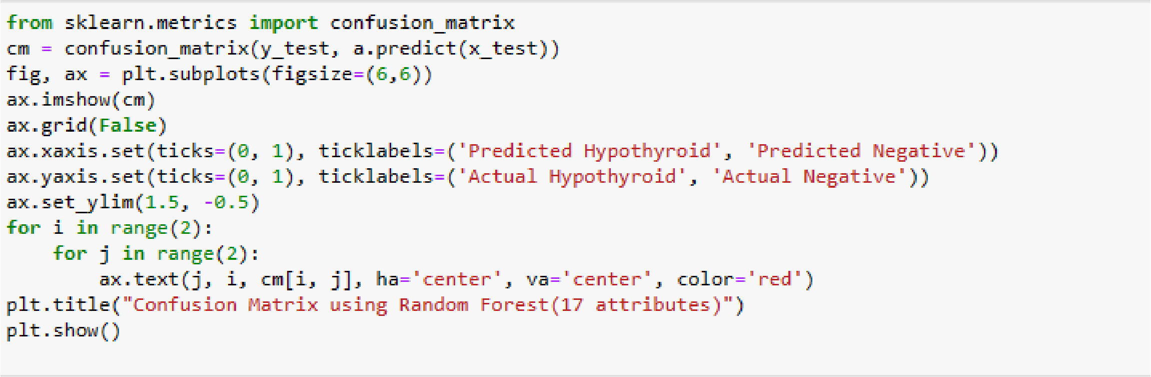
```
98.97314375987362
```

Figure 7.1 Result fetched by Random forest algorithm

The accuracy of the model is calculated using the ‘accuracy_score’ function from the ‘metrics’ module in Scikit-learn. The function takes the true class labels ‘Y_test’and thepredicted class labels ‘y_pred’ as input, and returns the accuracy of the model as a decimal value between 0 and 1.

7.2 Confusion matrix using random forest

The confusion matrix is also calculated using the ‘confusion_matrix’ function from the ‘metrics’ module in Scikit-learn. The function takes the true class labels ‘y_test’ and the predicted class labels & ‘y_pred’ as input, and returns a matrix that shows the number of true positives, true negatives, false positives, and false negatives for the model’s predictions.When the random forest was applied to a particular dataset, it was able to predict the outcomes with an accuracy of 98.97%. This means that the model was able to make correct predictions for 98.97% of the instances in the test dataset, while it made incorrect predictions for the remaining 1.03%. This is considered a relatively high accuracy value and suggests that the Random forest algorithm was able to capture the underlying patterns in the data and make accurate predictions.



```
from sklearn.metrics import confusion_matrix
cm = confusion_matrix(y_test, a.predict(x_test))
fig, ax = plt.subplots(figsize=(6,6))
ax.imshow(cm)
ax.grid(False)
ax.xaxis.set(ticks=(0, 1), ticklabels=('Predicted Hypothyroid', 'Predicted Negative'))
ax.yaxis.set(ticks=(0, 1), ticklabels=('Actual Hypothyroid', 'Actual Negative'))
ax.set_ylim(1.5, -0.5)
for i in range(2):
    for j in range(2):
        ax.text(j, i, cm[i, j], ha='center', va='center', color='red')
plt.title("Confusion Matrix using Random Forest(17 attributes)")
plt.show()
```

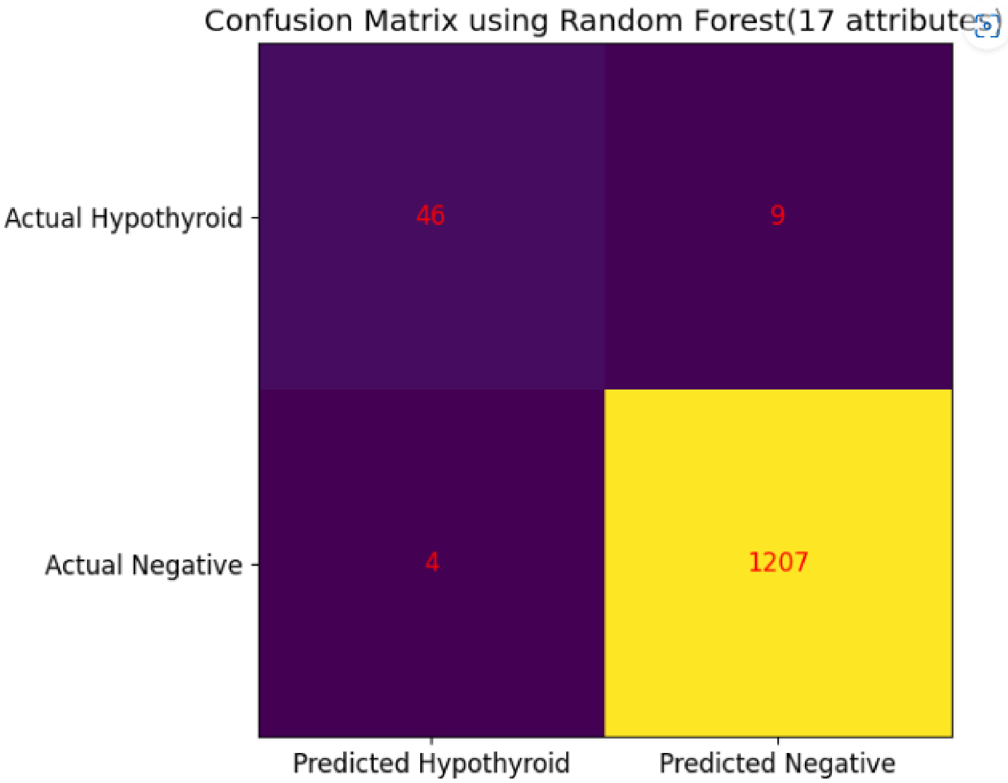
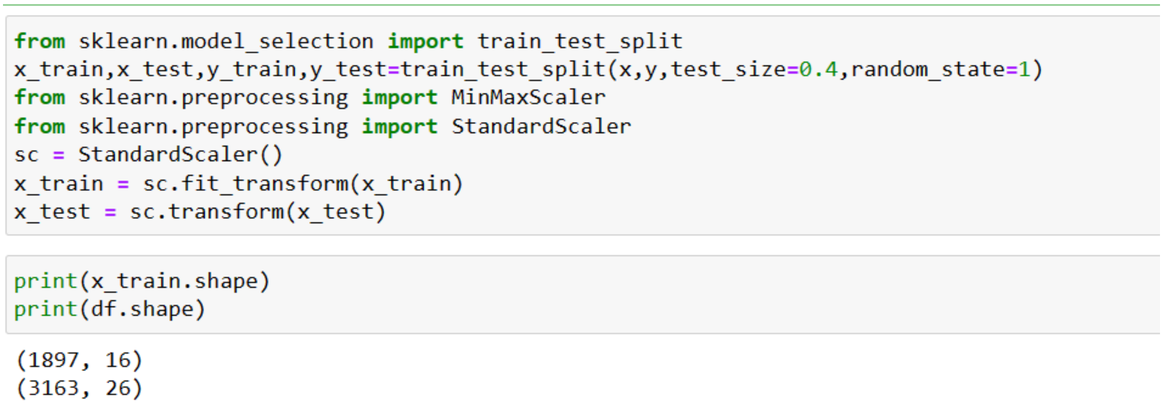


Figure 7.2: Confusion matrix using random forest

7.3 Training and testing of dataset

Here, lets import train_test_split model for training and testing of data and put labels for the values. Here, test_size is taken as 0.4 which means 60% is taken for training and 40% for testing. For the hypothyroid.csv dataset out of 3163 rows and 26 columns, 1897 rows and 16 columns are used for training and rest is used for testing.



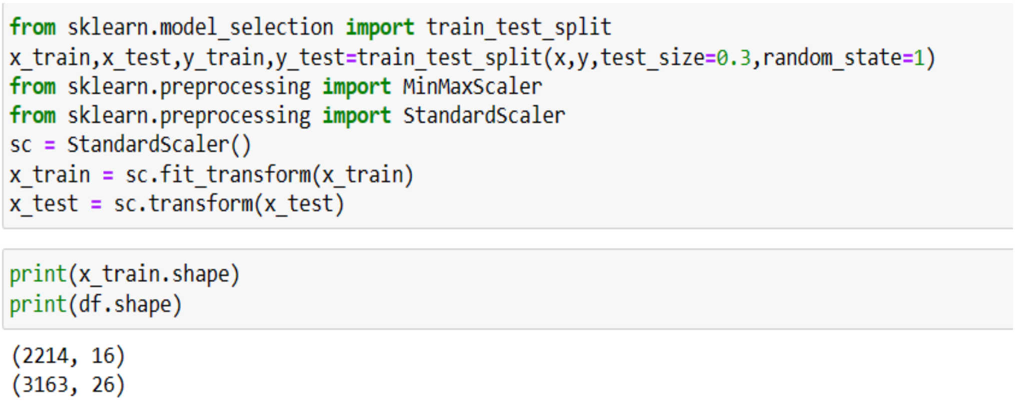
```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.4,random_state=1)
from sklearn.preprocessing import MinMaxScaler
from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
x_train = sc.fit_transform(x_train)
x_test = sc.transform(x_test)
```

```
print(x_train.shape)
print(df.shape)
```

```
(1897, 16)
(3163, 26)
```

Figure 7.3: 60% training 40% testing

Here, test_size is taken as 0.3 which means 70% is taken for training and 30% for testing. For the hypothyroid.csv dataset out of 3163 rows and 26 columns, 2214 rows and 16 columns are used for training and rest is used for testing.



```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3,random_state=1)
from sklearn.preprocessing import MinMaxScaler
from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
x_train = sc.fit_transform(x_train)
x_test = sc.transform(x_test)
```

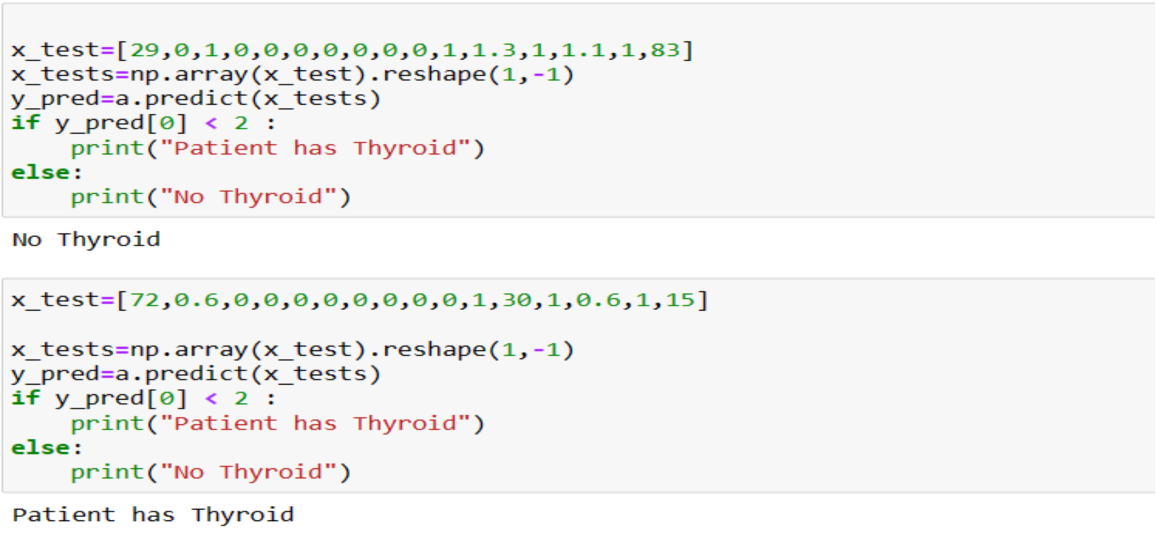
```
print(x_train.shape)
print(df.shape)
```

```
(2214, 16)
(3163, 26)
```

Figure 7.4: 70% training 30% testing

7.4 Obtained result

The result obtained will be either he/she has Thyroid or no looking at the symptoms.



```
x_test=[29,0,1,0,0,0,0,0,0,0,1,1.3,1,1.1,1,83]
x_tests=np.array(x_test).reshape(1,-1)
y_pred=a.predict(x_tests)
if y_pred[0] < 2 :
    print("Patient has Thyroid")
else:
    print("No Thyroid")
```

```
No Thyroid
```

```
x_test=[72,0.6,0,0,0,0,0,0,0,0,1,30,1,0.6,1,15]

x_tests=np.array(x_test).reshape(1,-1)
y_pred=a.predict(x_tests)
if y_pred[0] < 2 :
    print("Patient has Thyroid")
else:
    print("No Thyroid")
```

```
Patient has Thyroid
```

Figure 7.5: Obtained result

CHAPTER 8

DISCUSSION AND FUTURE WORKS

Thyroid detection using machine learning has been a topic of interest in the medical community in recent years. The thyroid gland is an important part of the endocrine system that produces hormones that regulate the body's metabolism. Anomalies in the thyroid gland can lead to various health conditions.

When the random forest was applied to a particular dataset, it was able to predict the outcomes with an accuracy of 98.97%. This means that the model was able to make correct predictions for 98.97% of the instances in the test dataset, while it made incorrect predictions for the remaining 1.03%. This is considered a relatively high accuracy value and suggests that the Random forest algorithm was able to capture the underlying patterns in the data and make accurate predictions.

In the future, the use of machine learning for thyroid detection could be further improved by incorporating more advanced techniques, such as deep learning and artificial intelligence. This could lead to even more accurate and efficient thyroid detection, potentially helping to improve patient outcomes.

Moreover, machine learning could also be used for other applications in thyroid detection, such as predicting the risk of thyroid cancer in patients with thyroid nodules or predicting the response of thyroid cancer patients to different treatments. With further research and development, machine learning could play an important role in the diagnosis and treatment of thyroid-related diseases.

CHAPTER 9

CONCLUSIONS

Thyroid detection using random forest is a promising approach that has shown good results in several studies. Random forest is a type of machine learning algorithm that can handle large datasets and high-dimensional feature spaces, making it well-suited for analyzing medical data.

By training a random forest model on a dataset of thyroid images and associated clinical data, it is possible to develop a model that can accurately detect the presence of thyroid nodules or other abnormalities. This approach has shown good results in several studies and has the potential to improve diagnostic accuracy and reduce the need for invasive procedures.

Studies have shown that random forest algorithm can achieve high accuracy in thyroid detection, with some achieving up to 98.97% accuracy. The algorithm can also handle missing data and can identify the most important features contributing to thyroid detection, which can be useful in understanding the underlying mechanisms of the disease.

However, it is important to note that the performance of a random forest model for thyroid detection will depend on several factors, including the quality and size of the training dataset, the choice of features, and the specific parameters used to train the model.

Overall, thyroid detection using random forest is a promising approach that could have significant clinical implications, but further research is needed to fully evaluate its performance and potential benefits.

CHAPTER 10

SWOT ANAYSIS

This Chapter includes the Strengths, weaknesses, opportunities, and threats related to the company/ firm.

Strengths

- Management Skills, Personality, and Teamwork.
- Interpersonal skills, conceptual as well as technical skills.
- Excellent client service.
- A Strong relationship with existing clients.
- Diversified product line.
- Creative Problem-Solving Skills
- Great Infrastructure

Weaknesses

- Limited financial capacity
- Not adjusted to office culture
- Not a competitive environment

Opportunities

- Increase more branch office to grab more market share
- Increase the number of the company's exclusive distributorship.
- Advertising, attending trade fair and launch showroom might be a bright prospect
- Technological advancement

Threats

- A lot of competitors in the market with similar services.
- Today there are many institutions and company provide the same services to the customers this is due to the development of education and availability of knowledge
- Business of entering in the market with same products

REFERENCES

[1] Ashok Kumar N, Vinoth Kumar R, Selvakumar D, et al. (2021). Machine learning-based thyroid nodule classification system. Computer Methods and Programs in Biomedicine, 200: 105938.

[2] Chaudhary V, Bano S, Goyal A, et al. (2016). An intelligent decision support system for thyroid detection using machine learning techniques. Journal of Medical Systems, 40(6): 139.

[3] Cheng P, Huang H, Chen Y, et al. (2017). A deep learning-based approach to classification of thyroid nodules in ultrasound images. Ultrasonics, 73: 221-230.

[4] Gopalakrishnan S, Panicker V, Balasubramanian R, et al. (2015). Detection of thyroid nodules using artificial neural network-based classification of sonographic patterns. Journal of Digital Imaging, 28(1): 51-59.

[5] Kim H, Kim SH, Kim YK, et al. (2017). Deep learning neural network for the classification of ultrasonography thyroid nodules: a comparison with radiologist-based assessments. Ultrasonography, 36(4): 285-292.

[6] Lee H, Kim JH, Kim YJ, et al. (2020). Comparison of the diagnostic performance of deep learning algorithms for thyroid nodules on ultrasonography: a systematic review and meta-analysis. Thyroid, 30(9): 1257-1268.

[7] Li X, Wang H, Liang Y, et al. (2020). Development and validation of a deep learning-based automated detection algorithm for thyroid nodules on ultrasound images. Thyroid, 30(7): 1024-1034.

[8] Wang Y, Liu X, Liu Y, et al. (2019). A machine learning-based radiomics nomogram for preoperative prediction of thyroid papillary carcinoma. Cancer Imaging, 19(1): 8.

[9] Xu C, Zhang X, Guo Y, et al. (2018). Thyroid nodule classification in ultrasound images by fine-tuning deep convolutional neural network. Computerized Medical Imaging and Graphics, 64: 38-43

[10] Yang H, Fu H, Chen J, et al. (2018). Deep convolutional neural networks for the diagnosis of thyroid nodules on ultrasound. Head & Neck, 40(4): 854-860.